



---

Morrison, G, Giovanis, Eleftherios, Pammolli, F and Riccaboni, M (2014) Border sensitive centrality in global patent citation networks. *Journal of Complex Networks*, 2 (4). pp. 518-536. ISSN 2051-1310

---

**Downloaded from:** <https://e-space.mmu.ac.uk/620967/>

**Version:** Accepted Version

**Publisher:** Oxford University Press

**DOI:** <https://doi.org/10.1093/comnet/cnu031>

Please cite the published version



## Border Sensitive Centrality in Global Patent Citation Networks

Journal:	<i>Journal of Complex Networks</i>
Manuscript ID:	CONNET-2014-025
Manuscript Type:	Article
Date Submitted by the Author:	24-Feb-2014
Complete List of Authors:	Morrison, Greg; IMT Lucca Inst. for Adv. Studies, Dept. of Economics Giovanis, Eleftherios; IMT Lucca Inst. for Adv. Studies, Dept. of Economics Pammolli, Fabio; IMT Lucca Inst. for Adv. Studies, Dept. of Economics Riccaboni, Massimo; IMT Lucca Inst. for Adv. Studies, Dept. of Economics
Subjects:	Social, socio-economic and political networks, Techno-social networks (i.e., online social networks, social networking sites, social media, Spatial networks, Multilayer networks

SCHOLARONE™  
Manuscripts

# Border Sensitive Centrality in Global Patent Citation Networks

Greg Morrison<sup>1,\*</sup>, Eleftherios Giovanis<sup>1</sup>, Fabio Pammolli<sup>1,2</sup>, and Massimo Riccaboni<sup>1,3,\*</sup>

<sup>1</sup> IMT Institute for Advanced Studies, Lucca Italy 55100

<sup>2</sup> Department of Economics Harvard University Cambridge MA, USA 02138

<sup>3</sup> Department of MSI, KU Leuven, Leuven, Belgium

\* to whom correspondence should be addressed

February 24, 2014

## Abstract

When resources are shared between networks in a multiplexed system, the importance of each node depends strongly on how collaborative or competitive each sub-network is. In this paper, we develop a new method of measuring centrality in the complex network of patent citations that can take political borders into account, where the national benefit of domestic citations relative to foreign citations can be controlled by a free parameter. We find that while some patent classes are of high importance both in the global and the domestic economy, there often exist patent classes in individual countries that are more central nationally than in global economy. We characterize the most important classes globally and domestically for six different nations, and describe their robustness to various perturbations to the model and to noise.

## 1 Introduction

The study of complex networks has been shown to be pertinent to a wide range of fields, including biology and medicine [1, 2, 3, 4], economics [5, 6], sociology [7, 8], and information science [9]. The topology of a network is of great importance in a variety of contexts [2, 3, 10, 4, 5], with significant attention paid to a variety of topologically defined node centralities [11, 6, 12, 13, 14], measuring the importance of an individual node relative to the rest of the network. Most centrality measures address the issue of importance for networks with edges of a single type, but new the emerging field of multiplex networks [15, 16, 17, 18, 19, 20] (or networks-of-networks) has begun to shed further light on the behavior of complex interacting systems by understanding the influence that one sub-system may have over another due to their interdependencies [19, 20, 21]. It is particularly

of interest to understand how influential a node is within its own network as well as within the entire multiplexed system [15, 21, 22], i.e. having a measure of centrality that incorporates the multiplexed topology of the global network.

In many contexts, networks-of-networks may be used to represent the interdependence between distinct groups of nodes that may either directly compete [21] or have an intrinsic preference for self-interaction. From the perspective of a sub-network  $A$ , sharing resources with another sub-network  $B$  may be considered detrimental (or at least not directly beneficial) from the perspective of  $A$ , whereas resources flowing in the opposite direction would be considered beneficial from the perspective of  $A$ . One example of such a system, which we focus on here, is the global network of national patent classes and the citations between them. Here the sub-networks represent national economies; nodes represent patent classes within those countries; and the information flow (the resource shared) models the technological spillovers [23, 24] between the various patent classes. Growth in one domestic patent class that is heavily cited by foreign patents may induce greater spillovers outside of the domestic economy, which may be considered detrimental to the domestic economy. While classes that are central to the global citation network will generally be those that are heavily cited [9], one may expect that domestically central classes are those that are simultaneously heavily cited primarily by *domestic* classes rather than foreign classes.

In this paper, an expansion on our previous work [22], we model the information flow originating in a specific national network as a random walk in the global citation network with an imposed loss term [14, 25] (a heightened discount rate [26, 27] in economic terms) depending on whether the information is propagating within the originating country or circulating outside of it. The loss term acts as a bias against border crossing, where information that benefits a foreign economy is assigned an elevated probability of becoming ‘lost’ (or incapable of benefiting the domestic economy in the future) [25, 28]. We define an expression for the centrality of domestic patent classes similar to PageRank [9, 14, 15] that varies continuously from a global measure of centrality (where there is no bias against information circulating in a foreign economy) to a domestic measure (which is maximally biased against border crossings) using a single parameter. We show that while heavily cited classes are often of importance both globally and domestically, there are many cases where smaller patent classes with a higher fraction of domestic citations are significantly more central domestically than they are globally. We show the model is robust to a wide range of parameter variations as well as random noise, and that the centralities of classes in larger economies are less likely to be perturbed by noise than those in smaller economies.

## 2 The patent citation network

Studying the structure of the patent citation network can provide insight into the dynamical processes of innovation that give rise to the filing of patents and the relationships between industries. Of particular interest are patents filed in the largest offices: the US Patent and Trademark Office (USPTO), the European Patent Office (EPO) and the Japanese Patent Office (JPO).. Each office has a method for categorizing the patents, and the EPO categorization into International Patent Classes (IPCs) is hierarchically organized [29]: each patent is assigned at least one 7-digit identifier with the first digit indicating the category of the patent (e.g. section H corresponds to electricity), the second and third digits the patent's class (e.g. H01 corresponds to 'basic electric elements'), the fourth digit a sub-class (e.g. H01C corresponds to 'Resistors'), and so on. By considering only triadic patents (related patents that cover the same claim in the EPO, USPTO, and JPO), we can apply the EPO hierarchy to aggregate similar patents that are filed in the USPTO or JPO. Studying the citation network of triadic patents has the additional advantage of focusing on 'important' patents: those that the inventor deemed worthy of applying for in three different offices. We construct the citation networks using raw data provided by the Organization for Economic Cooperation and Development (OECD), covering all citations between patents in the EPO and the World Intellectual Property Organization (WIPO) from 1978 to July 2013. We focus our attention primarily on the years 1987 to 2005, as the number of triadic patents decreases beyond 2005 due to a time lag in the granting of patents covering the same technology in all three offices (data not shown).

In order to aggregate the individual patent families into classes, we focus group patents together based on the first three digits of their IPC codes, proving  $n_d = 121$  patent classes that have at least one triadic patent filed in them since 1987. Patents with multiple IPC codes are treated as being members of all three-digit classes equally. The country of each patent is based on the address of the inventor(s), and patents with inventors in different countries are treated as being from both. Each country has different institutional expertise, and we therefore expect not all countries will be equally adept at all patent classes. We treat patent classes in each country as distinct from one another – class H01 in Germany is treated as distinct from H01 in Japan. Throughout this paper, we only consider the six largest economies as measured by the total number of triadic patents filed since 1987 as distinct sub-networks: the United States (US), Japan (JP), Germany (DE), France (FR), Great Britain (GB), and the Netherlands (NL). All other economies are grouped together as an external 'world' (WO) economy. These geographical-technological classes are treated as nodes in a set of directed graphs for year  $T$ , with the weighted edge from node  $i$  towards node  $j$  equal to the number of citations from class-country  $i$  in any year towards any patent in class-country  $j$  published in year  $T$  (discussed in greater detail in Sec. 3). The number of nodes in the network is therefore  $N = 7 \times n_d = 847$  nodes representing each class-country pair.

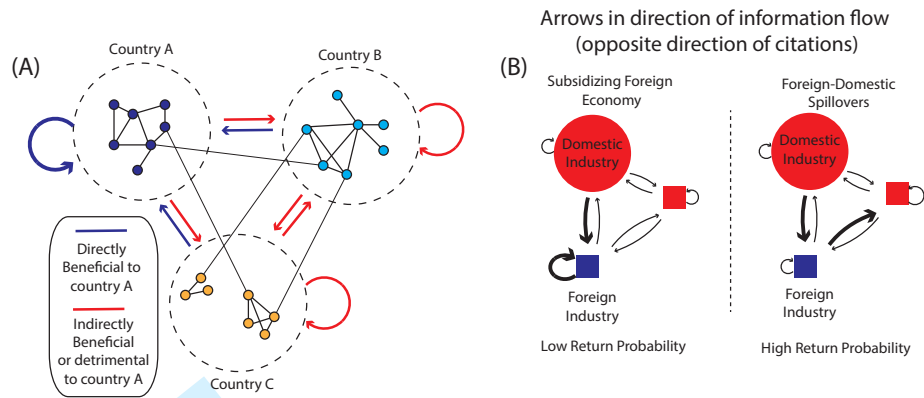


Figure 1: In (A), the patent classes for each country are denoted by nodes and citations between the classes by thin edges. Information flow from domestic classes towards foreign classes (indicated by the red arrows) may be expected to be less beneficial to the domestic economy than information remaining in the economy. This expectation does not imply that information lost to a foreign economy is guaranteed to be lost, as shown in (B). If a domestic industry is funded (the large circle) and is heavily cited by a foreign class, the information will likely move across the border (the thick arrow). The likelihood of the information returning depends on the network topology.

There are three levels of aggregation that are applied to the patent citation data in this approach: a geographical aggregation on the country level, a technological aggregation on the class level, and a temporal aggregation on the year level. All levels of aggregation have been applied in previous studies [30, 23, 5, 31] and are relevant to the triadic patent citation network. The IPC hierarchy is designed specifically to provide a meaningful relationship between classes, and political borders define significant regulatory, cultural, and industrial differences between locations. While different aggregations are possible (on the sector level [32] or on the level of 4-7 digit aggregation and regionally on the level of geography [30, 33]), we expect the three-digit coarse graining of the network to produce relevant information due to the diversity of classes (i.e. the representation is not too coarse) while remaining computationally approachable (i.e. an analysis can be performed quickly on a single computer). Temporal aggregation at a finer scale than yearly may be noisy, while grouping the citations into more coarse grained time windows (5 or 10 years for example) may make a meaningful study of the temporal behavior of the citation network difficult.

### 3 Random walks in innovation space

A single patent filed in year  $T$  may receive many citations from future patents, and in principle any of these citations may be viewed as a ‘spillover’ and act as a proxy for knowledge transfer [34]. We expect that the value of an innovation (represented by an individual patent) is blunted by the passage of time as new technologies are created [35], and it may therefore be reasonable to expect that a patent that is the ‘primary beneficiary’ of an innovation spillover (i.e. a patent that could not exist without the cited patent) is temporally close to the original patent. This spillover can be thought of as the first step in a random walk from the cited class to the citing class [9, 14], and so long as the structure of the global innovation network (quantified by the number of citations received by patents in class  $j$  filed in year  $T$  by patents in class  $i$  filed in any year,  $c_{i \rightarrow j}(T)$ ) does not change appreciably by the time the spillover occurs, we can treat the citation network as approximately static when determining the ‘primary beneficiary’ of future spillovers. If spillovers originating from a patent filed in year  $T$  occur more rapidly than large-scale structural changes to the economy, it is reasonable to approach the behavior of such spillovers as a random walk on the static citation network  $c_{i \rightarrow j}(T)$ .

The citation network  $c_{i \rightarrow j}(T)$  can be used to model innovation by recognizing that the probability of an innovation in class  $i$  being of benefit to class  $j$  should be a function of the fraction of citations from  $j$  to  $i$ ,  $f_{i \leftarrow j}$  (with the spillover moving in the opposite direction of the citations [36]). Spillovers are not expected to continue indefinitely (due partially to the blunting of the value of an innovation with time as mentioned above, and also to the possibility that an eventual spillover produces an uncited patent). We capture this expectation through a uniform probability of the walk ceasing at each step [14, 28],  $\epsilon_0$ , which acts as a discount parameter [26, 27] in an economic context. In our model of innovation spillovers as a random walk, the bare transition probability between classes is thus taken to be

$$p_{i \rightarrow j}^{bare} = (1 - \epsilon_0) f_{i \leftarrow j} = (1 - \epsilon_0) \frac{c_{i \leftarrow j}(T)}{\sum_k c_{i \leftarrow k}(T)}. \quad (1)$$

The probability of the random walk lasting at least  $k$  steps  $\sim e^{-k \log(1-\epsilon_0)}$ , so  $\epsilon_0$  sets the typical number of spillovers that are expected to occur before initial innovation becomes irrelevant. If the random walker is reinserted into the system when a loss occurs,  $\epsilon_0$  is equivalent to a teleportation probability commonly found in the PageRank centrality measure [14, 37, 25]. It is possible in such a model to define a steady state probability distribution, and we recover a PageRank measure of centrality:  $P_i = \sum_k [p_{k \rightarrow i}^{bare} + \epsilon_0/N] P_k$ , with  $N$  the number of nodes in the network and  $\epsilon_0 \equiv 0.15$  typically chosen. PageRank has already been highlighted as a useful centrality measure for citation networks in other contexts [9] and shown to be adaptable to other multiplexed networks [15].

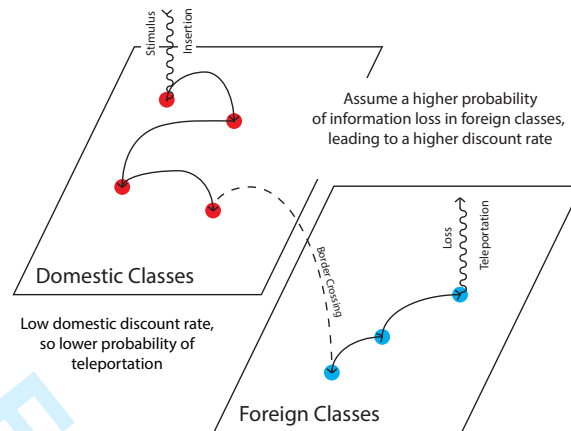


Figure 2: In our model, the walker is always inserted into the domestic economy, since the national government will only fund domestic classes directly. While propagating within the national economy a low discount rate is assumed, but once the information flows into the foreign economy the discount rate is elevated. The discounting of future value is reflected in the teleportation probability, with the random walk suddenly ending and the walker being re-inserted as a new investment in the domestic economy.

#### 4 Border sensitive centrality and asymmetric random walks

Central classes using PageRank will be those that heavily cite other globally central classes (i.e. classes  $i$  that have a large  $f_{k \leftarrow i}$  for other central classes  $k$ ) and, due to the near symmetry of the citation network, tend to also be heavily cited by important classes (data not shown). However, classes central to the domestic economy (rather than the global economy) will be simultaneously heavily cited *and* have a smaller fraction of citations originating from foreign economies (so that information does not flow as readily across the border). This can be modeled in the spirit of PageRank by introducing an asymmetric and border-dependent discount rate, equivalent to a nonuniform teleportation probability [28, 25] if the loss is coupled with a reinsertion. Our model permits four types of information flow: domestic-to-domestic (with an associated loss term  $\epsilon_{d \rightarrow d}$ ), foreign to domestic ( $\epsilon_{f \rightarrow d}$ ), domestic to foreign ( $\epsilon_{d \rightarrow f}$ ) and foreign to foreign ( $\epsilon_{f \rightarrow f}$ ). In this paper, we focus on a simplified model of information sharing where flow into the domestic economy is considered ‘good’ (so  $\epsilon_{d \rightarrow d} = \epsilon_{f \rightarrow d} \equiv \epsilon_0$  for the domestic-to-domestic and foreign-to-domestic discount rates), and information sharing benefiting foreign classes is considered ‘bad’ (so  $\epsilon_0 \leq \epsilon_{f \rightarrow f} = \epsilon_{d \rightarrow f} \leq 1$  for the domestic-to-foreign and foreign-to-foreign discount rates). We denote the elevation in the discount rate for border crossing by the parameter  $0 \leq \epsilon \leq 1$  and define  $\epsilon_{d \rightarrow f} = \epsilon_{f \rightarrow f} = 1 - (1 - \epsilon)(1 - \epsilon_0) \equiv \bar{\epsilon}$  (satisfying  $\epsilon_0 \leq \bar{\epsilon} \leq 1$ , so foreign spillovers are less beneficial than domestic spillovers).



We have defined the probability of reaching class  $i$  directly from class  $j$  without a loss occurring given their relationship in the global citation network (domestic-to-domestic, domestic-to-foreign, foreign-to-domestic, or foreign-to-foreign). To implement our asymmetric walk in the PageRank framework, we must re-insert the walker into the network once a loss occurs [14, 37, 25]. Since we are modeling the centrality from the perspective of a particular sub-network (a national economy), it is natural to re-insert the walker *only* within that subnetwork after loss rather than throughout the entire network. For example, we might reasonably expect that a national government might provide an external stimulus in a patent citation network only to domestic classes and never directly to foreign classes. Such a choice for a teleportation probability is equivalent to a topic-sensitive or personalized PageRank [11, 38]. If the insertion after loss is uniformly distributed amongst domestic classes, it is straightforward to show that the transition probabilities between classes becomes

$$p_{i \rightarrow j}^{d \rightarrow d}(\epsilon) = (1 - \epsilon_0) f_{i \leftarrow j} + \frac{\epsilon_0}{n_d} \left( f_{i \leftarrow D} + \frac{\bar{\epsilon}}{\epsilon_0} f_{i \leftarrow F} \right) \quad (2)$$

$$p_{k \rightarrow j}^{f \rightarrow d}(\epsilon) = (1 - \epsilon_0) f_{k \leftarrow j} + \frac{\epsilon_0}{n_d} \left( f_{k \leftarrow D} + \frac{\bar{\epsilon}}{\epsilon_0} f_{k \leftarrow F} \right) \quad (3)$$

$$p_{i \rightarrow l}^{d \rightarrow f}(\epsilon) = (1 - \bar{\epsilon}) f_{i \leftarrow l} \quad (4)$$

$$p_{k \rightarrow l}^{f \rightarrow f}(\epsilon) = (1 - \bar{\epsilon}) f_{k \leftarrow l} \quad (5)$$

where  $f_{i \leftarrow D} = \sum_{j \in \mathbf{D}} f_{i \leftarrow j}$  is the fraction of citations received by  $f$  from any class in the domestic economy (denoted  $\mathbf{D}$ ),  $f_{i \leftarrow F} = \sum_{l \in \mathbf{F}} f_{i \leftarrow l}$  is the fraction of citations from foreign classes (denoted  $\mathbf{F}$ ), and  $n_d$  is the number of classes in the domestic economy. To avoid traps, we assign a 100% teleportation probability for classes not simultaneously citing and cited by at least one class in the connected component by at least two citations. In an earlier version of this work [22], this restriction was relaxed to a constraint on the number citations (not received citations), which in some cases leads to small quantitative differences but no quantitative changes to the results.

All foreign and domestic classes can be assigned a domestic centrality ranking from the perspective of each country based on the steady state probability of finding the walker at that node. Defining the class' rank  $R_i^{(d)}(\epsilon)$  as the probability that the domestic class  $i$  is occupied at steady state during the random walk and  $R_k^{(f)}(\epsilon)$  as the steady state probability that foreign class  $k$  is occupied at steady state, it is straightforward to show that

$$R_i^{(d)}(\epsilon) = \sum_{j \in \mathbf{D}} p_{j \rightarrow i}^{d \rightarrow d}(\epsilon) R_j^{(d)}(\epsilon) + \sum_{l \in \mathbf{F}} p_{l \rightarrow i}^{f \rightarrow d}(\epsilon) R_l^{(f)}(\epsilon) \quad (6)$$

$$R_k^{(f)}(\epsilon) = \sum_{j \in \mathbf{D}} p_{j \rightarrow k}^{d \rightarrow f}(\epsilon) R_j^{(d)}(\epsilon) + \sum_{l \in \mathbf{F}} p_{l \rightarrow k}^{f \rightarrow f}(\epsilon) R_l^{(f)}(\epsilon). \quad (7)$$

The centrality of domestic classes can be determined without reference to the centrality of foreign classes through the eigenvalue equation

$$\mathbf{R}^{(d)}(\epsilon) = \left( \mathbf{P}^{d \rightarrow d}(\epsilon) + \mathbf{P}^{f \rightarrow d}(\epsilon)[\mathbf{1} - \mathbf{P}^{f \rightarrow f}(\epsilon)]^{-1} \mathbf{P}^{d \rightarrow f}(\epsilon) \right) \mathbf{R}^{(d)}(\epsilon), \quad (8)$$

where  $[\mathbf{P}^{x \rightarrow y}(\epsilon)]_{ij} = p_{j \rightarrow i}^{x \rightarrow y}(\epsilon)$  one of the four matrices of transition probabilities and  $\mathbf{R}^{(d)}(\epsilon)$  the vector of centralities of domestic classes. The domestic centrality  $\mathbf{R}^{(d)}(\epsilon)$  is thus an eigenvector of a matrix involving the domestic-to-domestic transition probabilities coupled with a convolution of the transition probabilities involving foreign economies. This eigenvalue problem has a variety of well known methods of solution [39], and we implement a simple power method to numerically determine  $\mathbf{R}^{(d)}$ . The eigenvector is normalized such that  $\sum_i R_i^{(d)}(\epsilon) = 1$  for all countries and all  $\epsilon$ .

The upper and lower bounds on the centrality of each country's innovation network will not in general be the same but rather depend strongly on the loss probability and the ratio of domestic-to-foreign citations in the former case and the skewness of the edge weight distribution in the latter case. In order to compare the variability of the importance of the same class between countries or for varying values of  $\epsilon$ , it will often be useful to examine the scaled centrality

$$\mu_i = \frac{R_i^{(d)} - \langle R^{(d)} \rangle}{\sigma_d} \quad (9)$$

where  $\langle R^{(d)} \rangle = n_d^{-1} \sum_i R_i \equiv n_d^{-1}$  the mean centrality of the nodes in the domestic economy (the latter equality due to our choice of normalization) and  $\sigma_d^2 = (n_d - 1)^{-1} \sum_i (R_i^{(d)} - n_d^{-1})^2$  is the variance of the centralities about the mean within the domestic economy. Large values of  $\mu_i$  correspond to very central classes of the domestic economy, and similar values of  $\mu_i$  in different countries indicate they are of similar significance in both countries, regardless of the upper and lower bounds on  $R_i$ .

## 5 Results for the Patent Citation Network

We apply our measure of centrality to the patent citation network described in Sec. 2 and choose a small value for the domestic loss probability  $\epsilon_0 = 0.01$  (the robustness of the model to this choice is discussed in Sec. 6). The behavior of important classes (those classes rated in the top 5 at either  $\epsilon = 0$  or  $\epsilon = 1$ ) are shown in Fig. 3 for 1995 and 2005. For comparison, we also show the scaled centrality defined in Eq. 9 in Fig. 4, which highlights the relative differences between these most important classes of the domestic economies. Only 21 classes appear in any country's top-5 list, as shown in the legends of Figs. 3 and 4, listed in descending order of the number of countries for which the class

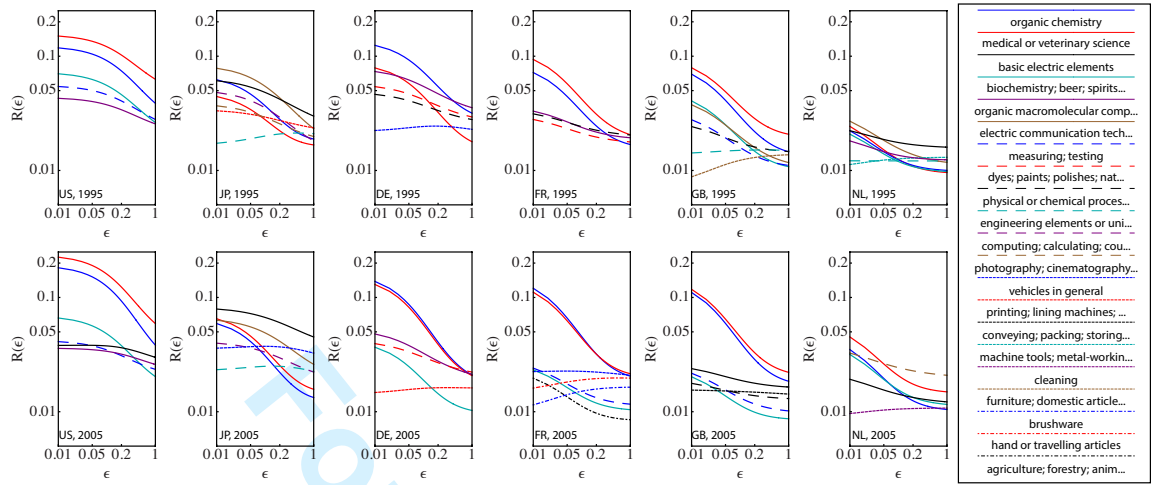


Figure 3: Shown are the centralities for the most important global or domestic classes for each country (those in the top-5 lists at either  $\epsilon = 0$  or  $\epsilon = 1$ ) as a function of  $\epsilon$  in 1995 (A) and 2005 (B) plotted on log-log axes. The decline observed in the most central classes is due to an increase in the overall loss probability at higher  $\epsilon$ .

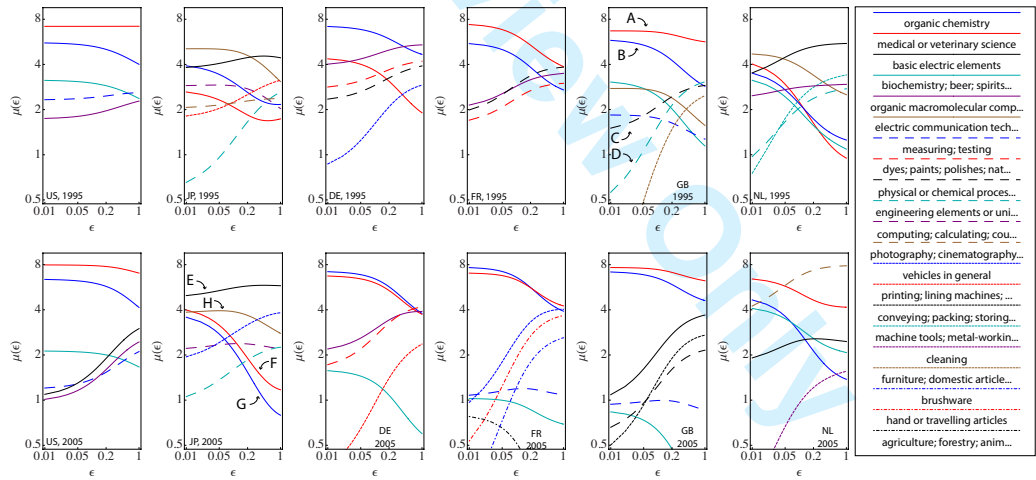


Figure 4: Shown is the behavior of scaled centrality as a function of  $\epsilon$  for the same classes as shown in Fig. 3, plotted on log-log axes. The additional labels A-D in the GB-1995 panel and E-G in the JP-2005 panel correspond to the same labels in Fig. 5 and Table 3 and are used as exemplars of the sometimes significant differences of globally and domestically central classes.

is central. ‘Organic Chemistry’ and ‘Medical or Veterinary Science’ are of overwhelming importance to most economies and are typically the two most central classes (with the prominent exception of Japan), but there is greater variety in the remaining classes in the remaining list of central classes for each nation.

We observe a decrease in the centrality of high-ranked classes at  $\epsilon = 0$  in most cases, due to the overall increase in the probability of loss at higher  $\epsilon$  in Fig. 3. Because the walker is reinserted randomly within the domestic economy, a higher overall probability of loss necessarily increases the centrality for unimportant classes (e.g. the centrality of a disconnected domestic node increases monotonically with  $\epsilon$ ) and therefore decreases the numerical value of  $R_i$  for important classes due to the normalization of the eigenvector. The *relative* change in centrality when comparing classes is entirely due to the topology of the citation network and is a signal of the difference between global and domestic centrality, most clearly indicated by the scaled centrality in Fig. 4. It is often the case that very central classes globally (at  $\epsilon = 0$ ) are of significantly lower importance domestically (at  $\epsilon = 1$ ). The primary exceptions to this trend are the ‘Organic Chemistry’ and ‘Medical or Veterinary Science’ in many countries (consistent with their place as most commonly central for all countries, discussed above). While there are multiple cases where these two classes do not depart from the top five for any  $\epsilon$ , it is particularly interesting to note that there is negligible change in  $\mu_i$  for ‘Medical or Veterinary Science’ in the US in 1995 (indicating that it’s importance to the US economy is essentially independent of the model parameters).

Additional details for the most central classes in Figs. 3 and 4 are listed in Table 1 for 1995 and Table 2 for 2005. While neither the extreme of a fully open economy (with no expected detriment to information sharing across borders and  $\epsilon = 0$ ) nor a fully protectionist economy (with no expected benefit from information sharing across borders and  $\epsilon = 1$ ) is likely to be optimal, the differences between these extremes in the tables illustrate the potential variance in the centrality for intermediate values of  $\epsilon$ . It is clear that the more heavily cited classes tend to be central globally (the left hand side of the tables), which is unsurprising since at  $\epsilon = 0$  the number of citations is the only determinant of centrality. At  $\epsilon = 1$  the influence of the borders becomes significant, as indicated by the increases in the columns for  $f_{i \leftarrow D}$  and  $f_{i \rightarrow D}$  (the fraction of forward and backward domestic citations). That is, classes that give or receive a high fraction of domestic citations tend to be more central domestically. While this bias towards domestic citations is expected from our model, our centrality does not simply return the domestic classes with the most citations. Globally central classes are sometimes large enough to remain central domestically (e.g. compare Japan’s ‘Electronic Communication Techniques’ with ‘Engineering Elements’, the 3<sup>rd</sup> and 4<sup>th</sup> most domestically central classes in 2005, Table2), and local topology sometimes having a strong effect (e.g. compare Japan’s ‘Basic Electric Elements’ with ‘Electronic Communication Techniques’, the 1<sup>st</sup> and 3<sup>rd</sup> most domestically central classes in 1995, Table1).

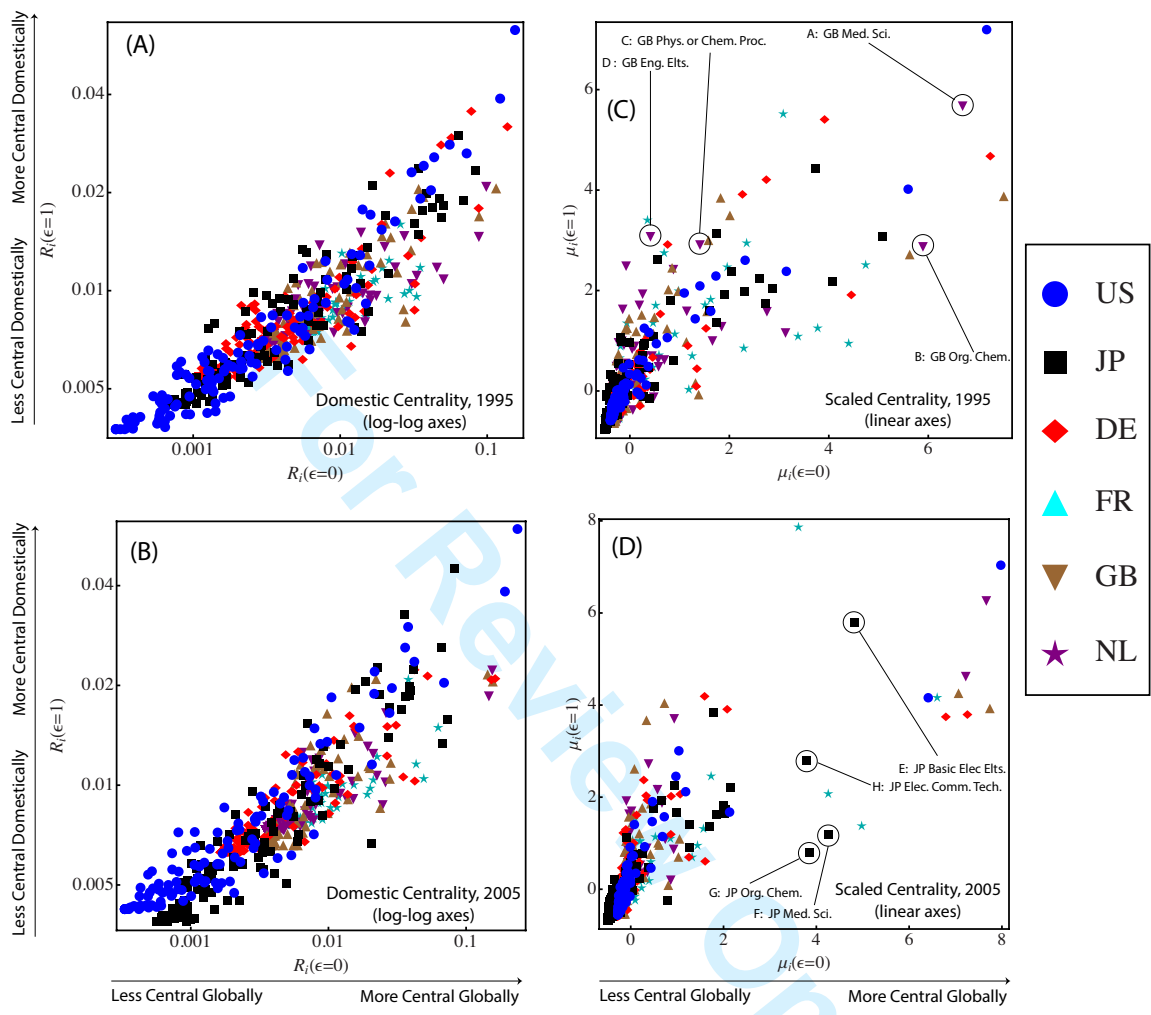


Figure 5: Shown is a comparison of the centralities for all classes and countries in 1995 (A) and 2005 (B) and the scaled centralities in (C) and (D) in the two extremes  $\epsilon = 0$  (measuring global centrality,  $x$ -axis) and  $\epsilon = 1$  (measuring domestic centrality,  $y$ -axis) with the domestic discount rate  $\epsilon_0 = 0.01$ . While there is a clear correlation between globally and domestically important classes in (A) and (B), the wide variance in the scaled centralities are evident in (C) and (D) even for very central classes. The labels A-H in the scaled centrality figures correspond to the same labels in Fig. 4 and Table 3 (with the labels A-D corresponding to the top four GB classes at  $\epsilon = 0$  in 1995 and the labels E-H corresponding to the top four JP domestic classes in 2005).

The centralities and scaled centralities for each class in each country are shown in Fig. 5 (A-B) and (C-D) respectively, and show the global centrality (at  $\epsilon = 0$ ) on the  $x$ -axes and domestic centrality (at  $\epsilon = 1$ ) on the  $y$ -axes. The global and domestic centralities are strongly related, with  $R_i(\epsilon = 1)$  typically falling within a factor of  $\sim 4$  of the global centrality of  $R_i(\epsilon = 1)$  over two decades of data. The scaled centrality more directly emphasizes the large impact on the rankings of very central classes due to the variation of a factor of  $\sim 4$  (particularly when shown with linear axes). We highlight a few classes for two countries in particular: the strong variations of the four most domestically central classes for GB in 1995 (marked with labels A-D in Fig. 5(C)) and of the four most globally central classes for JP in 2005 (marked with labels E-H in Fig. 5(D)). Table 3 shows the extreme variations in ranking that can occur as the loss probability  $\epsilon$  is varied using these classes as exemplars, with the sixteenth-highest global GB class at  $\epsilon = 0$  ranked second domestically (due to the high fraction of domestic citations, as listed in Table 1). Likewise, the second-highest class of global importance for JP in 2005 drops to 15<sup>th</sup> place as  $\epsilon$  increases due to its relatively low fraction of domestic citations (see Table 2). The variations illustrated in Fig. 5 are large in some cases, but an examination of the topology of the citation network makes it clear that our model captures a meaningful difference between globally and domestically important classes.

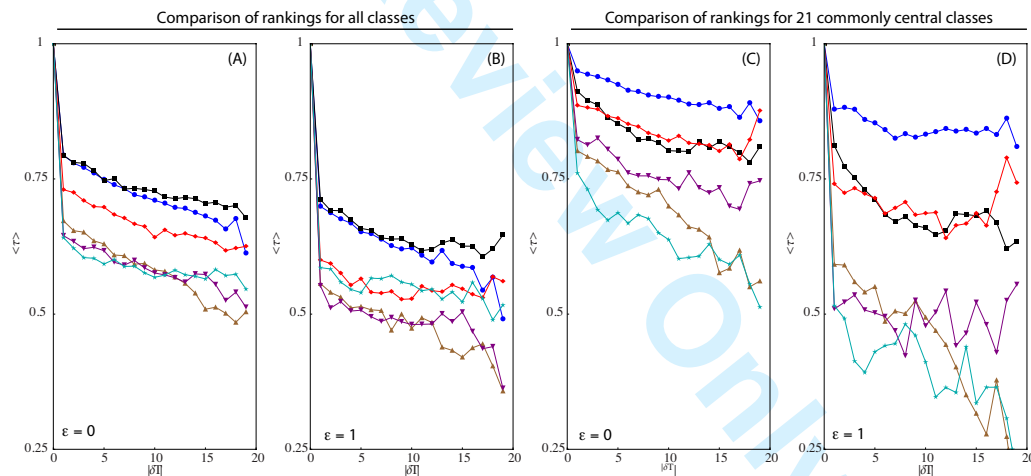


Figure 6: The Kendall-tau coefficient for the centralities of all classes at  $\epsilon = 0$  (A) and  $\epsilon = 1$  (B), and for the 21 major classes listed in Figs. 3 -4 and Tables 1 - 2 at  $\epsilon = 0$  (C) and  $\epsilon = 1$  (D). The sharp initial decline seen in (A-B) are greatly reduced in (C), indicating that globally central classes are slow to change. The initial decline of domestically central classes may be either very slow or very fast depending on the country, shown in (D).

We find the most central classes in the US and Japan are predominantly the same in 1995



and 2005 (over 66% of the members of the top-5 classes are found in both years), but there is a drastic restructuring observed for both France and the Netherlands (with below 40% of the top-5 classes the same in both years). This suggests that the centrality of classes in smaller economies may be less stable than those in larger economies as time progresses, which is confirmed by examining the Kendall-tau coefficient [40] (measuring the fraction of pairs of values in each list that have the same ordering) between centralities in different years for each country. In Fig. 6(A-B), the average of the Kendall-tau coefficient for each pair of years  $T_1$  and  $T_2$  is shown as a function of the time interval between them,  $|\delta T| = |T_1 - T_2|$ . Regardless of the value of  $\epsilon$ , Fig. 6(A-B) shows a steep drop by about 20-40% at  $|\delta T| = 1$  followed by a slow decline on the scale of decades for rankings for  $|\delta T| > 1$ . The initial drop at  $|\delta T| = 1$  is due primarily to changes in classes with low centrality as can be seen by restricting the Kendal-tau average to only the 21 important classes that are found in the top-5 for any country (those listed in Figs. 3-4 and Tables 1-2). The restricted average still shows a similar slow decline for global centralities in Fig. 6(C) with  $|\delta T| > 1$ , but with a greatly reduced initial decline at  $|\delta T| = 1$  indicating a smaller change in the relative ordering of these more central classes. For the domestic centralities (at  $\epsilon = 1$ ) in Fig. 6(D), the sharp decline at  $|\delta T| = 1$  is only reduced for the larger economies of US, JP, and DE. Smaller economies retain their large initial decline in the 21 most central classes and fluctuate more heavily as  $|\delta T|$  increases.

## 6 Robustness: Parameter Variation and Noise

It is valuable to understand the robustness of the model to variations in the parameters or aggregation and to statistical noise. In Fig. 7, we compare the centralities for two different values of the domestic loss parameter  $\epsilon_0$  using the correlation coefficient and Kendall- $\tau$  (determining the fraction of pairs of rankings that differ between the two orderings [40]) as a function of  $\epsilon$ . We compare the values of  $\epsilon_0 = 0.01$ , the loss probability considered throughout the paper, with the commonly used PageRank teleportation probability of  $\epsilon_0 = 0.15$ . While these loss probabilities differ by an order of magnitude, the agreement between centrality measures is very high: the correlation coefficient  $\rho$  above  $\sim 0.9$  and Kendall  $\tau$  above 0.85 for  $\epsilon = 0$  and improves steadily as  $\epsilon$  increases. This increase is due to the rarity of teleportation events at  $\epsilon = 0$  compared to  $\epsilon = 1$  (where in the latter case border crossings boost the overall probability of teleportation) and indicates that the domestic centrality should be considered at least as robust as the well-accepted topic-sensitive or personalized [11, 38] PageRank centralities to variations in the model parameters.

We also examine the effect of changing the temporal aggregation of the citation network by replacing the citations between classes received in year  $T$ ,  $c_{i \leftarrow j}(T)$ , with that of a three-year total  $c_{i \leftarrow j}^*(T) = c_{i \leftarrow j}(T-1) + c_{i \leftarrow j}(T) + c_{i \leftarrow j}(T+1)$ . A comparison between the centralities for all classes and countries for the 1-year and 3-year aggregations is shown in Fig. 8 for

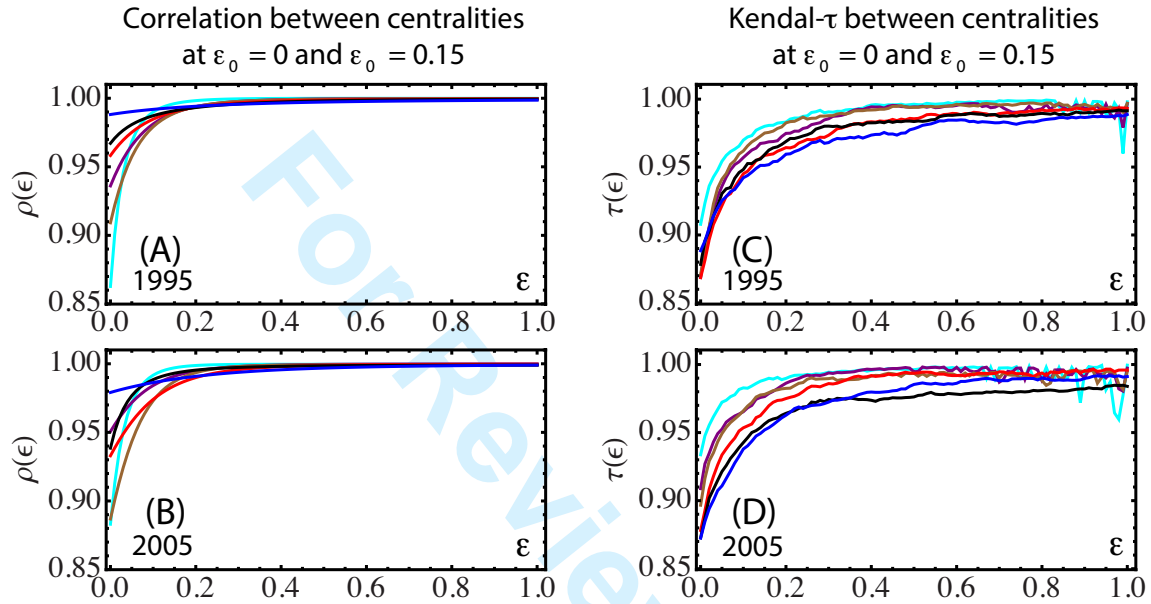


Figure 7: As a robustness check on the method, we quantify the similarity between the centrality vector  $\mathbf{R}^{(d)}(\epsilon)$  for two different teleportation probabilities:  $\epsilon_0 = 0.01$  (the very low value used throughout this paper) and  $\epsilon_0 = 0.15$  (the traditional PageRank value). The correlation coefficient and the Kendall- $\tau$  coefficients are shown in (A-B) and (C-D) respectively for varying  $\epsilon$  with USA in blue, Japan in black, Germany in red, France in brown, Great Britain in purple, and the Netherlands in cyan. (A) and (C) show 1995, (B) and (D) show 2005. There is a good agreement between the centralities for the global measure (near  $\epsilon = 0$ ) which improves with increasing  $\epsilon$ , approaching near-perfect agreement at  $\epsilon \sim 1$ .



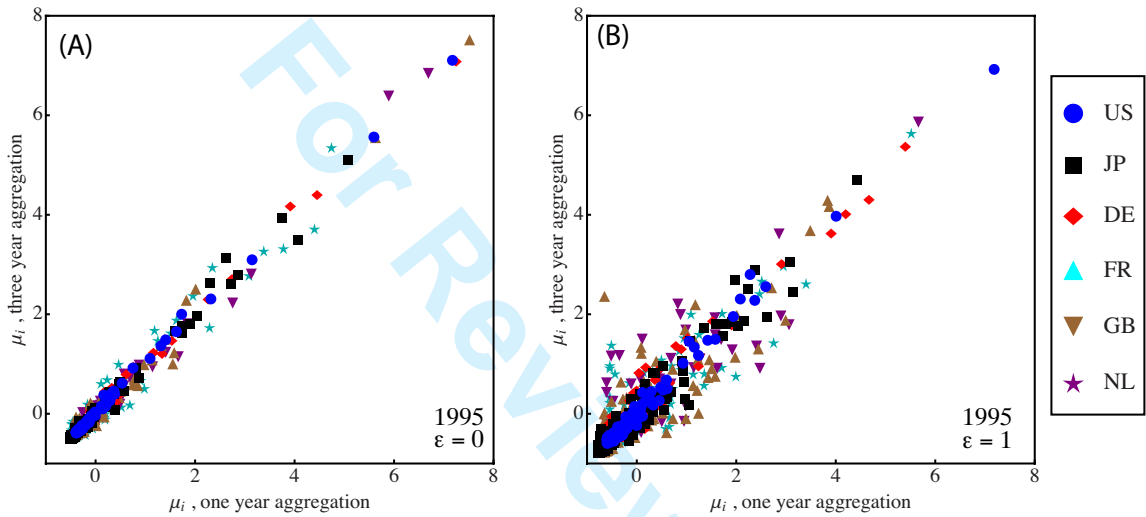


Figure 8: Shown is a comparison of centralities as different levels of temporal aggregation. The  $x$  axes correspond to node centrality, computed from networks based on citations to classes in 1995 only, while the  $y$  axis is correspond to centralities for a temporal aggregation of 1994, 1995, and 1996 (a three year window). The scaled global (A) and domestic (B) centralities are highly correlated for central classes (those with large  $\mu_i$ ), but the domestic centrality shows greater variability for non-central classes (those with  $\mu_i \lesssim 0$ ).

$T = 1995$ , and that there is a strong correlation between the centralities for both citation networks. Consistent with the results of Fig. 6, temporal variations have a greater effect on domestic centrality than global centrality (as evidenced by the greater variance in Fig. 8(B)), and the centrality of unimportant classes tends to be more strongly effected than those of highly central classes.

As a final check on the robustness of the model, we simulate the effect of noise in the number of citations between classes by randomly rewiring the observed network and re-computing the centralities of the classes. The rewiring is diagrammed in Fig. 9 (A): a pair of edges connecting four distinct nodes are randomly chosen and the locations of their endpoints are swapped. This rewiring procedure guarantees that the degree of each node is preserved since no edges are added or removed (just exchanged) but does not preserve the fraction of domestic or foreign citations. We therefore expect the rewire to have a small effect on the global centrality (at  $\epsilon = 0$ ) and a more significant effect on domestic centrality (at  $\epsilon = 1$ ) where the nationality of the citations are relevant. This expectation is confirmed in Fig. 9(B-C), which shows the average root mean square deviation (RMSD) difference between the rewired centralities and the unmodified centralities in 1995:  $\langle \delta R(\epsilon) \rangle = n_d^{-1} \sum_i [n_r^{-1} \sum_r (R_{i,r}(\epsilon) - R_i(\epsilon))^2]^{1/2}$ , with  $R_{i,r}(\epsilon)$  the centrality of class  $i$  in simulation  $r$  and  $n_r = 50$  the number of rewire simulations performed. Fig. 9 confirms a number of expected behaviors. First, the average RMSD increases monotonically with  $\epsilon$ , meaning that global centrality is more robust to rewires that preserve degree than domestic centrality. We also see that increasing the fraction of rewires increases the overall RMSD for all countries (compare the axes of Figs. 9(B) and (C)). Finally, the centralities of classes in smaller economies like NL, GB, and FR tend to be more susceptible to fluctuations in the citation network than the larger economies of US, JP, and DE, consistent with the behavior seen in Fig. 6.

## 7 Conclusions

Centrality has long been recognized as a useful concept in determining the importance of individual nodes in complex networks, and patent citation networks are known to be amenable to such analysis as well [41, 6]. In the global network of interdependent (and competing) national innovation systems, the importance of a patent class in the context of a domestic economy may not coincide with its importance in a global sense. In this paper, we define a new measure of centrality that takes into account a national preference for keeping innovation spillovers within the domestic economy rather than allowing it to flow into a foreign economy. We construct a model of innovation spillovers as a random walk and introduce the preference for domestic spillovers as an asymmetric loss term that depends on the political borders. We use this model to define a modified PageRank measure of centrality that provides insights into the patent classes that are important *domestically*,

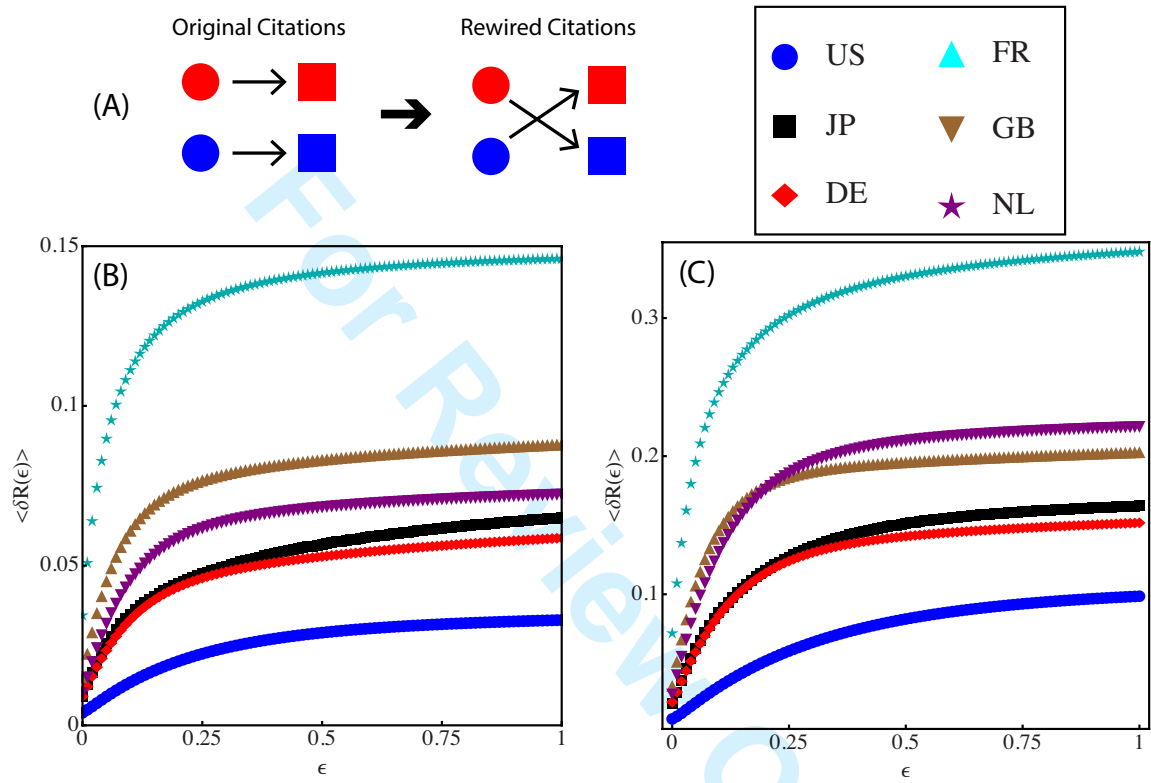


Figure 9: Shown is the robustness of our measure of centrality under a degree-preserving random rewiring, schematically diagrammed in (A). (B) shows the RMSD difference  $\langle \delta R(\epsilon) \rangle$  as a function of  $\epsilon$  after rewiring 5% of the edges in 1995, while (C) shows the same for a 25% rewire fraction. The monotonic increase in the RMSD indicates that domestic centrality, which depends most strongly on the details of cross-border citations, is far more susceptible to the perturbation than is global centrality.

in terms of a single parameter  $\epsilon$  (the relative increase in the probability of loss as information crosses a border). Our model produces a wealth of information about the structure of national economies in a global network, identifying patent classes of importance both globally and from a more self-interested domestic perspective. We expect this approach can be fruitfully applied to a variety of other multiplexed systems where sub-networks are both interdependent and competing [21] to determine the most central nodes globally or in each sub-network.

## 8 Acknowledgments

The authors would like to thank Andrea Morescalchi for numerous helpful discussions and an anonymous reviewer of a previous version of this work. We acknowledge funding from the CRISIS Lab project and from the Multiplex FP7 project (Foundational Research on MULTilevel comPLEX networks and systems).

## References

- [1] K. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. Barabasi. The human disease network. *Proc. Natl. Acad. Sci.*, 104:8685, 2007.
- [2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:5594, 2002.
- [3] K. Yan, G. Fang, N. Bhardwaj, R. P. Alexander, and M. Gerstein. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc. Natl. Acad. Sci.*, 107:9186, 2010.
- [4] D. Meunier, R. Lambiotte, and E. T. Bullmore. Modular and hierarchically modular organization of brain networks. *Front. in Neurosci.*, 4:1, 2010.
- [5] C. A. Hidalgo, B. Klinger, A.-L. Barabasi, and R. Hausmann. The product space conditions the development of nations. *Science*, 317:482, 2007.
- [6] F. Blochl, F. Theis, F. Vega-Redondo, and E. Fisher. Vertex centralities in input-output networks reveal the structure of modern economies. *Phys. Rev. E*, 83:046127, 2011.
- [7] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter. Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.*, 53:526, 2011.

- [8] N. Eagle, M. Macy, and R. Claxton. Network diversity and economic development. *Science*, 328:1029, 2010.
- [9] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding scientific gems with Google's PageRank algorithm. *J. Infometrics*, 1:8, 2007.
- [10] G. Morrison and L. Mahadevan. Asymmetric network connectivity using weighted harmonic averages. *Europhys. Lett.*, 93:40002, 2011.
- [11] T. H. Haveliwala. Topic sensitive PageRank: A context sensitive ranking algorithm for web search. *IEEE Trans. Knowl. Data Eng.*, 15:784, 2003.
- [12] P. Bonachich. Power and centrality: A family of measures. *Amer. J. Soc.*, 92:1170, 1987.
- [13] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Soc. Net.*, 27:39, 2005.
- [14] M. Franceschet. PageRank: standing on the shoulders of giants. *Com. ACM*, 54:92, 2011.
- [15] A. Halu, R. J. Mondragon, P. Panzarasa, and G. Bianconi. Multiplex pageRank. *PLoS One*, 8:e78293, 2013.
- [16] V. Nicosia, G. Bianconi, V. Latora, and M. Barthélemy. Growing multiplex networks. *Phys. Rev. Lett.*, 111:058701, 2013.
- [17] S. Gomez et. al. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.*, 110:028701, 2013.
- [18] J. Gomez-Gardenes, I. Reinares, A. Arenas, and L. M. Floria. Evolution of cooperation in multiplex networks. *Scientific Reports*, 2:620, 2012.
- [19] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464:1025, 2010.
- [20] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Staneley. Robustness of a network of networks. *Phys. Rev. Lett.*, 107:195701, 2011.
- [21] J. Aguirre, D. Papo, and J. Buldu. Successful strategies for competing networks. *Nature Physics*, 9:230, 2013.
- [22] G. Morrison, E. Giovannis, M. Riccaboni, and F. Pammolli. Border sensitive centralities in patent citation networks using asymmetric random walks. *Intl. Conf. on Signal-Image Tech. and Internet-Based Systems*, page 546, 2013.
- [23] D. T. Coe and E. Helpman. International R&D spillovers. *Eur. Econ. Rev.*, 39:859, 1995.

- [24] P. Erdi, K. Makovi, Z. Somogyvari, K. Strandburg, J. Tobochnik, P. Volf, and L. Zalai. Prediction of emerging technologies based on analysis of the US patent citation network. *Scienom.*, 95:225, 2013.
- [25] R. Rossi and D. Gleich. Dynamic PageRank using evolving teleportation. In *Algorithms and Models for the Web Graph*, page 126. Springer, 2012.
- [26] D. Pearce, B. Groom, C. Hepburn, and P. Koundouri. Valuing the future: recent advances in social discounting. *World Econ.*, 4:121, 2003.
- [27] K. Arrow, M. Cropper, C. Gollier, B. Groom, G. Heal, R. Newell, W. Nordhaus, R. Pindyck, W. Pizer, P. Portney, T. Sterner, R. S. J. Tol, and M. Weitzman. Determining benefits and costs for future generations. *Science*, 341:349, 2013.
- [28] P. G. Constantine and D. F. Gleich. Random alpha PageRank. *Int. Math.*, 6:189, 2009.
- [29] A full list of EPO categories can be found at <http://web2.wipo.int/ipcpub>.
- [30] J. Alcacer and W. Chung. Location strategies and knowledge spillovers. *Manag. Sci.*, 53:760, 2007.
- [31] G. Papaconstantinou, N. Sakurai, and A. Wyckoff. Domestic and international product-embodied R&D diffusion. *Research Policy*, 27:301, 1998.
- [32] U. Schmoch, F. LOaville, P. Patel, and R. Frietsch. *Linking Technology Areas to Industrial Sectors*. European Commission, 2003.
- [33] A. Chessa, A. Morescalchi, F. Pammolli, O. Penner, A. M. Petersen, and M. Riccaboni. Is Europe evolving towards an integrated research area? *Science*, 339:650, 2013.
- [34] A. B. Jaffe, M. Trajtenberg, and P. M. Romer. *Patents, citations, and innovations: A window on the knowledge economy*. MIT Press, 2002.
- [35] A. B. Jaffe and M. Trajtenberg. Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proc. Natl. Acad. Sci.*, 93:12671, 1996.
- [36] A. O. Zhirov, O. V. Zhirov, and D. L. Shepelyansky. Two dimensional ranking of wikipedia articles. *Eur. Phys. J. B*, 77:523, 2010.
- [37] D. F. Gleich, A. D. Flaxman, P. G. Constantine, and A. Gunawardana. Tracking the random surfer: Empirically measured teleportation parameters in PageRank. *Proc. 19th Intl Conf. on WWW*, page 381, 2010.
- [38] D. Fogaras and B. Racz. *Algorithms and Models for the Web-Graph*. 2004.
- [39] P. Berkhin. A survey on pagerank methods. *Internet Math.*, 2:73, 2005.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

[40] M. Kendall. A new measure of rank correlation. *Biometrika*, 30:81, 1938.

[41] B. Gress. Properties of the USPTO patent citation network: 1963-2002. *World Patent Info.*, 32:3, 2010.

For Review Only

Table 1: Details of the top-five ranked classes in 1995 are listed for  $\epsilon = 0$  (left) and  $\epsilon = 1$  (right), with  $\epsilon_0 = 0.01$  throughout.  $\mu_i = (R_i - \langle R \rangle) / \sigma_R$  is the scaled centrality of each class,  $C$  is the total number of citations received by that class for patents filed in 2005 (excluding those from the WO category of countries not individually considered), and  $f_{i \leftarrow D}$  and  $f_{i \rightarrow D}$  are the fraction of citations incoming from and outgoing towards the domestic economy, respectively. For  $\epsilon = 0$  the more commonly patented classes tend to be highly ranked, while for  $\epsilon = 1$  there is a bias towards patent classes that cite and are cited by other domestic classes.

US, year=1995, $\epsilon_0 = 0.01$ . $\langle C \rangle = 663.479$ , $C_{max} = 11894$ .									
$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.17	11.9k	75%	74%	medical or veterinary scien...	7.18	11.9k	75%	74%	medical or veterinary scien...
5.6	10.5k	73%	69%	organic chemistry	4.01	10.5k	73%	69%	organic chemistry
3.15	6.3k	77%	73%	biochemistry; beer; spirits...	2.59	4.9k	74%	67%	measuring; testing
2.32	4.9k	74%	67%	measuring; testing	2.37	6.3k	77%	73%	biochemistry; beer; spirits...
1.73	4.3k	61%	55%	organic macromolecular comp...	2.28	4.3k	61%	55%	organic macromolecular comp...
JP, year=1995, $\epsilon_0 = 0.01$ . $\langle C \rangle = 338.975$ , $C_{max} = 3537$ .									
$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
5.07	3.5k	64%	59%	electric communication tech...	4.43	2.9k	63%	55%	basic electric elements
4.07	2.9k	50%	51%	organic chemistry	3.14	1.5k	75%	72%	printing; lining machines; ...
3.74	2.9k	63%	55%	basic electric elements	3.07	3.5k	64%	59%	electric communication tech...
2.87	2.3k	57%	54%	computing; calculating; cou...	2.61	0.4k	58%	67%	engineering elements or uni...
2.72	2.1k	44%	41%	medical or veterinary scien...	2.37	1.7k	73%	63%	photography; cinematography...
DE, year=1995, $\epsilon_0 = 0.01$ . $\langle C \rangle = 226.967$ , $C_{max} = 4143$ .									
$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.25	4.1k	46%	59%	organic chemistry	5.4	2.9k	46%	56%	organic macromolecular comp...
4.45	2.7k	39%	49%	medical or veterinary scien...	4.67	4.1k	46%	59%	organic chemistry
3.91	2.9k	46%	56%	organic macromolecular comp...	4.2	1.8k	52%	65%	dyes; paints; polishes; nat...
2.74	1.8k	52%	65%	dyes; paints; polishes; nat...	3.91	1.4k	46%	59%	physical or chemical proces...
2.26	1.4k	46%	59%	physical or chemical proces...	2.91	0.5k	52%	60%	vehicles in general
FR, year=1995, $\epsilon_0 = 0.01$ . $\langle C \rangle = 101.702$ , $C_{max} = 1852$ .									
$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.52	1.9k	38%	45%	medical or veterinary scien...	3.87	1.9k	38%	45%	medical or veterinary scien...
5.63	1.6k	40%	43%	organic chemistry	3.84	0.6k	48%	51%	physical or chemical proces...
2.01	0.8k	36%	42%	organic macromolecular comp...	3.49	0.8k	36%	42%	organic macromolecular comp...
1.82	0.6k	48%	51%	physical or chemical proces...	2.99	0.4k	33%	55%	dyes; paints; polishes; nat...
1.58	0.4k	33%	55%	dyes; paints; polishes; nat...	2.71	1.6k	40%	43%	organic chemistry
GB, year=1995, $\epsilon_0 = 0.01$ . $\langle C \rangle = 105.545$ , $C_{max} = 1833$ .									
$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
6.69	1.8k	43%	38%	medical or veterinary scien...	5.66	1.8k	43%	38%	medical or veterinary scien...
5.89	1.8k	40%	32%	organic chemistry	3.06	0.2k	49%	57%	engineering elements or uni...
3.13	1.1k	34%	33%	biochemistry; beer; spirits...	2.9	0.4k	33%	35%	physical or chemical proces...
2.76	0.9k	20%	29%	electric communication tech...	2.86	1.8k	40%	32%	organic chemistry
1.84	0.7k	34%	27%	measuring; testing	2.48	25	92%	43%	furniture; domestic article...
NL, year=1995, $\epsilon_0 = 0.01$ . $\langle C \rangle = 34.2479$ , $C_{max} = 371$ .									
$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
4.74	0.4k	20%	20%	electric communication tech...	5.52	0.2k	38%	28%	basic electric elements
4.4	0.3k	23%	22%	medical or veterinary scien...	3.4	18	83%	28%	machine tools; metal-workin...
3.78	0.3k	24%	15%	organic chemistry	2.95	0.3k	25%	28%	organic macromolecular comp...
3.38	0.2k	24%	16%	biochemistry; beer; spirits...	2.75	38	32%	35%	engineering elements or uni...
3.08	0.2k	38%	28%	basic electric elements	2.51	0.4k	20%	20%	electric communication tech...



Table 2: The same as in Table 1, but for classes in 2005.

US, year=2005,  $\epsilon_0 = 0.01$ .  $\langle C \rangle = 548.413$ ,  $C_{max} = 13635$ .

$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.98	13.6k	71%	69%	medical or veterinary scien...	7.03	13.6k	71%	69%	medical or veterinary scien...
6.41	12.k	67%	66%	organic chemistry	4.15	12.k	67%	66%	organic chemistry
2.13	5.7k	74%	65%	biochemistry; beer; spirits...	3.	3.k	58%	58%	basic electric elements
1.18	3.5k	76%	61%	measuring; testing	2.44	2.6k	64%	62%	organic macromolecular comp...
1.04	3.k	58%	58%	basic electric elements	2.11	3.5k	76%	61%	measuring; testing

JP, year=2005,  $\epsilon_0 = 0.01$ .  $\langle C \rangle = 375.653$ ,  $C_{max} = 4139$ .

$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
4.81	4.1k	71%	67%	basic electric elements	5.79	4.1k	71%	67%	basic electric elements
4.27	3.k	44%	47%	medical or veterinary scien...	3.83	1.5k	74%	77%	vehicles in general
3.84	3.k	46%	44%	organic chemistry	2.79	3.k	65%	74%	electric communication tech...
3.79	3.k	65%	74%	electric communication tech...	2.25	0.8k	74%	77%	engineering elements or uni...
2.15	2.k	58%	71%	computing; calculating; cou...	2.2	2.k	58%	71%	computing; calculating; cou...

DE, year=2005,  $\epsilon_0 = 0.01$ .  $\langle C \rangle = 180.14$ ,  $C_{max} = 4242$ .

$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.26	4.2k	41%	46%	organic chemistry	4.18	1.1k	49%	54%	dyes; paints; polishes; nat...
6.79	3.6k	35%	46%	medical or veterinary scien...	3.9	1.4k	50%	58%	organic macromolecular comp...
2.08	1.4k	50%	58%	organic macromolecular comp...	3.79	4.2k	41%	46%	organic chemistry
1.6	1.1k	27%	38%	biochemistry; beer; spirits...	3.74	3.6k	35%	46%	medical or veterinary scien...
1.59	1.1k	49%	54%	dyes; paints; polishes; nat...	2.36	0.3k	59%	77%	printing; lining machines; ...

FR, year=2005,  $\epsilon_0 = 0.01$ .  $\langle C \rangle = 89.6116$ ,  $C_{max} = 2486$ .

$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.74	2.5k	30%	43%	organic chemistry	4.25	2.4k	34%	43%	medical or veterinary scien...
7.06	2.4k	34%	43%	medical or veterinary scien...	4.03	0.2k	25%	42%	vehicles in general
1.05	0.4k	26%	44%	measuring; testing	3.92	2.5k	30%	43%	organic chemistry
1.03	0.6k	20%	34%	biochemistry; beer; spirits...	3.66	50	82%	80%	hand or travelling articles
0.81	0.5k	16%	45%	agriculture; forestry; anim...	2.61	23	100%	82%	brushware

GB, year=2005,  $\epsilon_0 = 0.01$ .  $\langle C \rangle = 89.7603$ ,  $C_{max} = 2582$ .

$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
7.66	2.6k	38%	32%	medical or veterinary scien...	6.25	2.6k	38%	32%	medical or veterinary scien...
7.22	2.4k	40%	32%	organic chemistry	4.61	2.4k	40%	32%	organic chemistry
0.94	0.4k	39%	38%	basic electric elements	3.69	0.4k	39%	38%	basic electric elements
0.92	0.6k	33%	21%	measuring; testing	2.71	65	32%	48%	conveying; packing; storing...
0.86	0.8k	44%	19%	biochemistry; beer; spirits...	2.15	0.4k	35%	37%	physical or chemical proces...

NL, year=2005,  $\epsilon_0 = 0.01$ .  $\langle C \rangle = 43.6198$ ,  $C_{max} = 731$ .

$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class	$\mu$	$C$	$f_{i \leftarrow D}$	$f_{i \rightarrow D}$	class
6.6	0.7k	15%	28%	medical or veterinary scien...	7.87	0.3k	42%	55%	photography; cinematography...
4.98	0.7k	15%	26%	organic chemistry	4.16	0.7k	15%	28%	medical or veterinary scien...
4.26	0.6k	16%	33%	biochemistry; beer; spirits...	2.46	0.3k	28%	24%	basic electric elements
3.61	0.3k	42%	55%	photography; cinematography...	2.08	0.6k	16%	33%	biochemistry; beer; spirits...
1.73	0.3k	28%	24%	basic electric elements	1.56	15	53%	76%	cleaning

Label	Class	global rank	domestic rank
A	GB Med. Sci.	1	1
B	GB Organic Chem.	2	4
C	GB Phys. or Chem. Proc.	7	3
D	GB Engineering Elts.	16	2
E	JP Basic Elec. Elts.	1	1
F	JP Med. Sci.	2	15
G	JP Organic Chem.	3	18
H	JP Elec. Comm. Techniques	4	3

Table 3: Details of the highlighted top-5 classes in Fig. 4 and Fig. 5 and their global ( $\epsilon = 0$ ) and domestic ( $\epsilon = 1$ ) ranks. A-D correspond to the four most domestically central GB classes in 1995, and (E-H) correspond to the four most globally central JP classes. Classes that are globally unimportant (like GB Engineering Elements) may be very central domestically, and likewise globally central classes (like JP's Organic Chem.) may have a low domestic centrality.